This is a manuscript of an article that was accepted in January 2018 for publication in the *Journal of Hydrology*.

The final publication is available at https://doi.org/10.1016/j.jhydrol.2018.01.044

Modeling Urban Coastal Flood Severity from Crowd-Sourced Flood Reports Using Poisson Regression and Random Forest

J. M. Sadler [1], J. L. Goodall [1], M. M. Morsy [1,2], K. Spencer [3]

[1]Dept. of Civil and Environmental Engineering, Univ. of Virginia, 351 McCormick Rd., P.O. Box 400742, Charlottesville, VA 22904

[2]Irrigation and Hydraulics Dept., Cairo University, P.O. Box 12211, Giza 12614, Egypt

[3]Geographic Information Systems Department, City of Norfolk, 401 Monticello Ave., Norfolk, VA 23510

Corresponding author: Jonathan Goodall, `goodall@virginia.edu`

Abstract

Sea level rise has already caused more frequent and severe coastal flooding and this trend will likely continue. Flood prediction is an essential part of a coastal city's capacity to adapt to and mitigate this growing problem. Complex coastal urban hydrological systems however, do not always lend themselves easily to physically-based flood prediction approaches. This paper presents a method for using a data-driven approach to estimate flood severity in an urban coastal setting using crowd-sourced data, a non-traditional but growing data source, along with environmental observation data. Two data-driven models, Poisson regression and Random Forest regression, are trained to predict the number of flood reports per storm event as a proxy for flood severity, given extensive environmental data (i.e., rainfall, tide, groundwater table level, and wind conditions) as input. The method is demonstrated using data from Norfolk, Virginia USA from September 2010 - October 2016. Quality-controlled, crowd-sourced street flooding reports ranging from 1-159 per storm event for 45 storm events are used to train and evaluate the models. Random Forest performed better than Poisson regression at predicting the number of flood reports and had a lower false negative rate. From the Random Forest model, total cumulative rainfall was by far the most dominant input variable in predicting flood severity, followed by low tide and lower low tide. These methods serve as a first step toward using data-driven methods for spatially and temporally detailed coastal urban flood prediction.

1 Introduction

Flooding in low-lying, coastal cities has become more common in recent years due to climate change and relative sea level rise [*Sweet and Park*, 2014]. Relative sea level is expected to rise substantially [*Vermeer and Rahmstorf*, 2009; *Church et al.*, 2001], which will worsen the problem of flooding in coastal cities. Flooding in coastal cities is caused by large, life-threatening, high-return period events such as Hurricanes Harvey and Irma whose flooding recently severely affected coastal cities in Texas and Florida USA, respectively. Additionally, many coastal cities have low-relief terrain and low elevation making stormwater drainage problematic. This can make coastal cities susceptible to flooding from smaller, low-return period events such as severe thunderstorms. The long-term effects of legacy engineering decisions can further add to an urban city's flood risk (e.g., the use of non-engineered fill used to reclaim streams which causes higher than average subsidence rates [*Turner*, 2004]).

The ability to accurately predict flooding allows decision makers to proactively mitigate the effects of flooding [*Zevenbergen et al.*, 2008], which is key to a city's resilience to natural hazards [*Godschalk*, 2003]. Accurate flood prediction allows decision makers to maximize safety in the case of large events, and minimize infrastructure damage and social and economic disruption in the case of smaller events. Accurate flood prediction also allows cyber-physical (or smart) stormwater systems to perform optimally, further mitigating the effects of flooding [*Kerkez et al.*, 2016].

Modeling and predicting flooding in urban coastal environments can be challenging. Urban coastal floods are influenced by a combination of different environmental, geographic, and human-related factors [*Gallien et al.*, 2014]. Environmental factors that contribute to coastal flooding include rainfall, wind, tide levels, and ground water table levels. Geographic factors such as elevation, soil properties, proximity to the coast, and the land use and land cover of the drainage area can influence whether a given location experiences flooding. In urban settings human-related factors including built stormwater infrastructure and the condition of that infrastructure, which is often underground, also play a role in the location and severity of flooding. For example, clogged stormwater inlets and undersized stormwater pipes can increase the chance and severity of flooding. High tidal levels can inundate stormwater outfalls rendering them ineffective at draining stormwater to the ocean, a condition which will become more frequent with sea level rise. The need to accurately represent such systems and their changing conditions further adds to the complexity of urban flood modeling.

Urban coastal flood events can be modeled using physically-based 1D [*Mark et al.*, 2004] or 2D models [*Mignot et al.*, 2006; *Hunter et al.*, 2008; *Bates et al.*, 2005; *Smith et al.*, 2011; *Gallien et al.*, 2014]. However, the simplified representations of reality used in physically-based models can be a limitation given the combination of variables and their interactions, and the complexity of the physical environment. Two-dimensional, hydrodynamic models make fewer simplifications compared to 1D models, however, this comes at a larger computational cost [*Leandro et al.*, 2009] which makes executing, and especially calibrating, a 2D model difficult [*Caviedes-Voullième et al.*, 2012].

Another modeling approach shown to be effective in many fields [*Yang et al.*, 2017a] including hydrology [*Solomatine and Ostfeld*, 2008] is data-driven modeling. Data-driven models detect patterns in the data to map model inputs to model outputs without attempting to simulate the physical processes [*Solomatine and Ostfeld*, 2008]. Thus, the relationship between the inputs and outputs is not assumed, as in physically-based models, but learned. While physical processes are not directly simulated using data-driven models, understanding of physical processes usually influences the selection of input variables used to predict the output variable [*Booker and Woods*, 2014].

The recent increase in availability of earth observation data, coupled with advances in machine learning algorithms, have expanded the possibilities and use of data-driven modeling in hydrology. Machine learning algorithms have been used extensively in hydrology for applications such as predicting reservoir operations [*Yang et al.*, 2016], soil mineral weathering [*Povak et al.*, 2014], streamflow [*Yang et al.*, 2017b; *Solomatine and Price*, 2004; *Wang et al.*, 2009], groundwater potential [*Naghibi et al.*, 2017], and groundwater level [*Sahoo et al.*, 2017]. Data-driven and machine learning algorithms in flooding applications specifically have been used by *Tehrany et al.* [2013], *Wang et al.* [2015], and *Tien Bui et al.* [2016] who predicted areas susceptible to flooding, *Adamovic et al.* [2016], who modeled flash flooding on a regional scale, and *Solomatine and Price* [2004], who predicted streamflow for flood forecasting. Despite the expanded use of data-driven models in hydrology, few studies have used data-driven methods to model flooding within coastal urban environments. The closest work may be the statistical analysis of tidal records in the United States to estimate the amount of time that coastal cities have experienced flooding in the past several decades and project flooding in the coming decades [*Ezer and Atkinson*, 2014; *Sweet and Park*, 2014; *Moftakhari et al.*, 2015; *Ray and Foster*, 2016].

The objective of this study is to use data-driven modeling to predict flooding severity for a given storm in an urban coastal setting. Crowd-sourced flood reports recorded during flood events are used for model training and are considered a proxy variable for flood severity. Although a more objective measure of flood severity is preferred to the number of flood reports (e.g., flood inundation depth and duration throughout the study domain), often such data is not available. Relevant environmental data (rainfall, tide levels, water table level, wind speed and direction) will be used as inputs to the model.

A data-driven approach is appropriate for this application due to the complexity of modeling urban coastal flooding, as discussed above, which makes using a physical model difficult. This paper will investigate and compare two different data-driven models, Poisson regression and Random Forest regression. Poisson regression is a generalized linear model and was selected because it is commonly used to model rare events [*D'Unger et al.*, 1998] and a flood report, while increasing in occurrence, can still be considered a rare event. Random Forest was selected due to its wide use as a machine learning algorithm in hydrology applications [*Yang et al.*, 2016; *Wang et al.*, 2015; *Loos and Elsenbeer*, 2011] and other fields [e.g., *Mutanga et al.*, 2012; *Svetnik et al.*, 2003].

The data-driven approach will be applied in Norfolk, Virginia USA. Norfolk and the surrounding Hampton Roads region is one of the most vulnerable metropolitan centers to coastal flooding in the USA [*Fears*, 2012]. Since 2010, the City has collected quality-controlled, crowd-sourced street flooding reports for 45 storm events. In this study, the two data-driven models, Poisson regression and Random Forest regression, will be trained to predict the number of street flood reports per storm event, given the rainfall, tidal, water table, and wind characteristics of the storm event. The models will be evaluated and compared using primarily the root mean squared error (RMSE) and mean absolute error (MAE) between the predicted number of street flood reports and the actual number of street flood reports.

This is a first step toward the use of data-driven approaches in urban coastal flood modeling. Additionally, although *Gaitan et al.* [2016] employed exploratory methods to glean information from open spatial data, weather data, and user reports, the use of crowd-sourced data in the training and evaluation of data-driven predictive models for urban flood modeling has not been demonstrated or discussed thoroughly in the literature. This is relevant currently as multiple platforms now exist for collecting crowd-sourced information regarding urban flooding [*Le Coz et al.*, 2016] and it can be expected that, due to the nearly universal use of internet connected devices, crowd-sourced data will continue to grow in volume. It is also anticipated that the results of the model will shed light on the relative importance of different environmental factors in predicting coastal flooding, another subject that has been given little attention in previous literature regarding urban coastal flooding.

The remainder of this paper will proceed as follows. First, background will be given describing the study area, the model input and output data, and an introduction to the data-driven models used. Next, the methods are presented describing the preparation of the data for the models and how the data-driven models were applied and evaluated. The model results are then presented and discussed, and finally conclusions are given.

2 Study Area, Data, and Model Background

2.1 Study Area and Street Flooding Record

Norfolk, Virginia USA, shown in Figure 1, is an ideal study area for this research considering its vulnerability to flooding, its economic and military importance, and the availability of quality-controlled crowd-sourced data regarding flood occurrences for the city. Norfolk is one of the most vulnerable cities to coastal flooding in the USA due largely to land subsidence rates causing Norfolk and the surrounding area to experience relative sea level rise at a rate faster than the global average [*Kleinosky et al.*, 2006]. As home to the largest terminal of the Port of Virginia, the 3rd most used port on the East Coast of the US [*The Port of Virginia*, 2016], Norfolk plays an important role in the economy of Virginia and the surrounding states. The world's largest naval base, Naval Station Norfolk, is also within Norfolk, making the flooding risks in the area important to US national security [*Broder*, 2009].

An important factor in selecting the study area was the availability of crowd-sourced flood record data. A record of reported flooded street locations has been kept in Norfolk starting with Hurricane Nicole on 30 September 2010, shown in Figure 1. This is a unique dataset because often, observational data from flooding events is a limiting factor in creating useful flood models [*Smith et al.*, 2011]. Because such data is often sparse, photographs of flooded locations and personal interviews have been used out of necessity in the calibration and verification of flood models [*Smith et al.*, 2011]. Even satellite imagery has been used to estimate flooding extents [*Ireland et al.*, 2015], but is less useful as a street-scale flood record in an urban setting due to its coarse spatial resolution.

2.2 Description of Model Input and Output Data

The objective of this study is to develop a model capable of predicting flooding severity resulting from a given storm event based on the environmental conditions of that event. The environmental condition input data for the model consisted of rainfall, water table level, wind, and tide level observations. These were obtained from the Hampton Roads Sanitation District (HRSD) and the US National Oceanic and Atmospheric Administration (NOAA). From HRSD, rainfall, water table elevation, and wind direction and wind speed data were obtained. The rainfall observations were on a 15-minute time scale, and the water table elevations and wind data were on a 2-minute time scale. From NOAA, 6-minute water elevations and daily high and low tides recorded at the Sewells Point [*NOAA*, 2017a] and Money

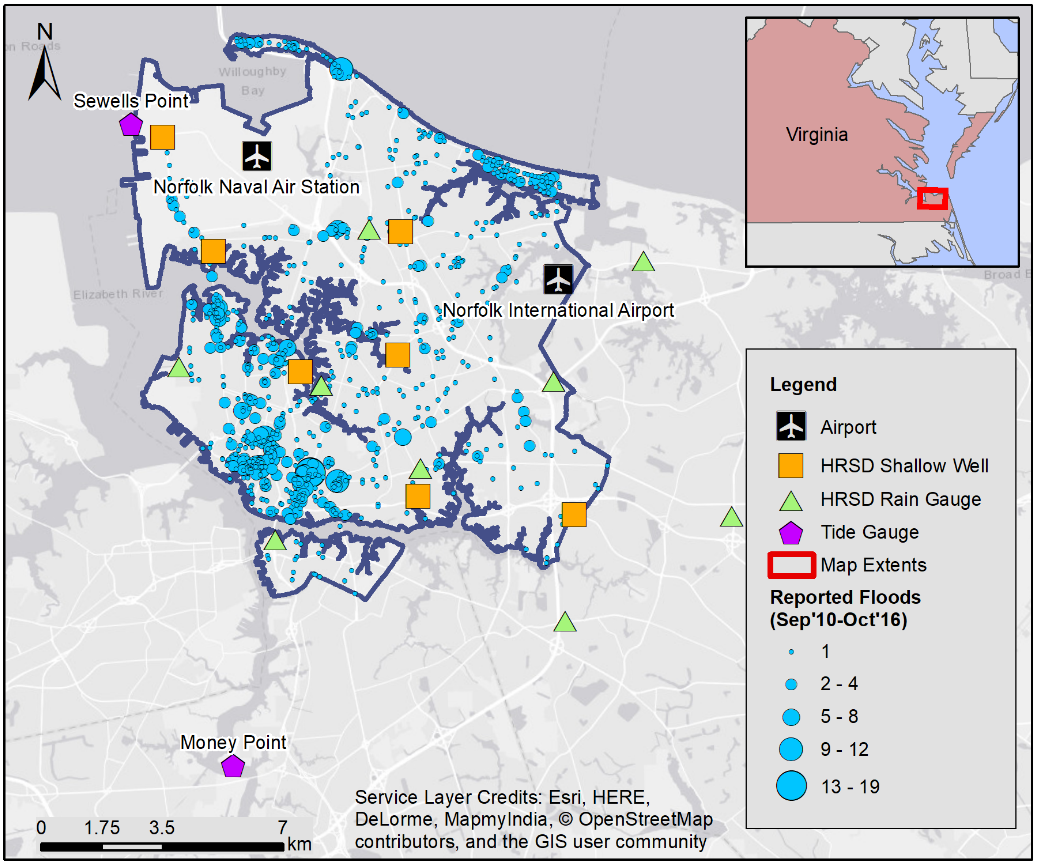

Figure 1. Study area: Norfolk, Virginia USA

Point [*NOAA*, 2017b] tide gauges were obtained. Wind speed, wind gust, and wind direction data recorded at the Money Point station at 6-minute time intervals were used as well. Daily rainfall and wind data collected at two airports in the study area, Norfolk International Airport [*NOAA*, 2017c] and Norfolk Naval Air Station [*NOAA*, 2017d], were also obtained from NOAA. The rain gauge, water table, wind, and tide gauge stations and airports are shown in Figure 1. All of the raw data together consisted of more than 15 million observations. To keep the time series data organized, a simplified version of the Consortium of Universities for the Advancement of Hydrologic Sciences Incorporated (CUAHSI) Observations Data Model [*Horsburgh et al.*, 2009] was implemented in a sqlite database.

The target data used for the model training and evaluation were the number of crowd-sourced flooded location reports per storm event from September 2010 to October 2016. The flood reports were made and catalogued using the City's custom System to Track, Organize, Record, and Map (STORM). STORM is used by the City to record, and catalogue impacts from storms on the City's infrastructure (e.g., downed powerlines, damaged trees) and STORM data are viewable online at http://gisapp1.norfolk.gov/stormmap. For most of the study period, only City of Norfolk staff were able to make reports in STORM however, in the Spring of 2016 the mobile application used for reporting was made available to the public. Reports made by the general public underwent an approval process by City staff.

The two categories of STORM reports used for model training in this study were "flooded street" and "flooded underpass". A total of 45 storm events (listed in Table 1) were reported to have caused at least one flooded street or flooded underpass in the period of record. The number of reported floods per event ranged from 1 to 159. Figure 2 shows a box plot of the street flood reports per event. Eight of the events were hurricanes; the rest were unnamed or given generic names by city workers. The number of flood reports made from the top six storm events were much larger than the number of reports made from the other 39 storm events. These larger events are marked as points and labeled in Figure 2.

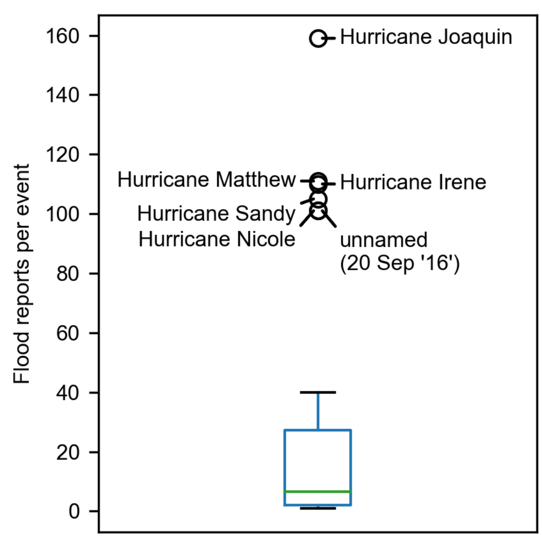

Figure 2. Summary plot of reported floods per event in Norfolk, VA Sep. 2010 - Oct. 2016

2.3 Model Alternatives

2.3.1 Poisson Regression

Poisson regression is a generalized linear model (GLM) commonly used to model rare event, count data. Applications of Poisson regression include modeling crime rate [*Osgood*,

Table 1. Events recorded to have caused flooding in Norfolk Sep. 2010 - Oct. 2016

Event Date	Event Name	Flood Reports
29 Sep 2015	Hurricane Joaquin	159
05 Oct 2016	Hurricane Matthew	111
27 Aug 2011	Hurricane Irene	110
28 Oct 2012	Hurricane Sandy	105
20 Sep 2016	unnamed	101
30 Sep 2010	Hurricane Nicole	101
02 Sep 2016	Hurricane Hermine	40
10 Jul 2014	Thunderstorms	39
09 Oct 2013	Heavy Rain	36
16 May 2014	Heavy Rain	35
08 Sep 2014	Rainy Monday	31
20 Jan 2016	January Winter Weather	26
24 Jul 2014	unnamed	18
24 Sep 2015	Noreaster	16
19 Sep 2016	Heavy Rain	11
02 Mar 2015	unnamed	10
11 Jul 2015	Thunderstorm	10
19 Jul 2016	Thunderstorm	9
25 Feb 2016	unnamed	8
03 Jul 2014	Hurricane Arthur	8
31 Jul 2016	Thunderstorm	8
02 Jul 2015	unnamed	7
15 Jan 2016	unnamed	6
03 Jun 2016	Severe Weather - 6/5	6
04 Sep 2014	Thunderstorm	5
19 Jun 2014	Thunderstorms	5
01 Feb 2016	unnamed	5
23 Nov 2014	unnamed	4
13 Sep 2014	Saturday Storm	3
30 Dec 2015	Heavy Rainfall	3
09 Jul 2014	Thunderstorms	2
25 Jul 2016	Bernie (Training)	2
10 Jun 2016	unnamed	2
29 Sep 2014	unnamed	2
16 Dec 2010	Snow	2
24 Feb 2016	February 24th Storm	1
17 Nov 2014	Storm	1
30 Oct 2015	unnamed	1
20 Jul 2016	unnamed	1
17 Sep 2015	unnamed	1
02 Sep 2015	unnamed	1
18 Aug 2014	unnamed	1
24 Sep 2014	Heavy Rain	1
09 Jun 2014	unnamed	1

2000], disease incidence [*Frome and Checkoway*, 1985], and manufacturing defects [*Lambert*, 1992]. *Morrison and Smith* [2002] and *Viglione et al.* [2014] used Poisson distributions to model the arrival time and occurrence of flood peaks, respectively.

There are two main assumptions made when using Poisson regression. The first is that the response variable (number of flood reports in this case) follows a Poisson distribution

$$P = e^{-\lambda}\frac{\lambda^k}{k!} \tag{1}$$

where P is the probability that k number of events will occur per interval of time and λ is the event rate. The second major assumption when using Poisson regression is that the variance and the mean of the response variable are equal. Thus, the probability distribution (eq. 1) can be specified by only one parameter, λ [*Coxe et al.*, 2009].

In Poisson regression, the mean parameter, λ, is defined by the log-linear function

$$\lambda = e^{-x_i \beta} \tag{2}$$

where x_i is a vector of input values for time i and β is a corresponding vector of model parameters, which is optimized during training [*Cameron and Trivedi*, 1998].

2.3.2 Random Forest

Random forest, developed by *Breiman* [2001], is an ensemble machine learning algorithm which uses a large number of classification or regression trees (CART) to make a prediction [*Breiman et al.*, 1984]. The response variable in this case, the number of flood reports per event, is modeled using regression, therefore the Random Forest model is an ensemble of regression trees. In the training of a regression tree, rules based on the response variable are developed to divide observations until the resulting predictions have a minimum amount of node impurity. Node impurity for regression trees, as defined by *Breiman et al.* [1984], is the sum of the squared deviations between the predicted and actual value [*Loh*, 2011]. The regression tree's rules are a collection of linear divisions of the observation data that, together, create a non-linear decision surface.

One of the main problems of regression trees is that they are prone to overfitting to the training data and thus perform poorly when given unseen data [*Murphy*, 2012]. Random Forest is an approach that attempts to address this weakness. When an individual regression tree is trained in the Random Forest algorithm, a portion of the input records and predictor variables are randomly selected as input to the training. This process is repeated for the number of regression trees specified by the modeler, thus creating a group of regression trees, each trained on a randomly selected subset of the records and input variables. This group of regression trees constitutes a Random Forest model. The prediction made by a Random Forest regression model is the average of the predictions made by each individual regression tree. The random selection of input records and variables in the training of the individual regression trees creates variety in the weak learners, thus avoiding overfitting of the model to the training data.

Beyond the actual predictive capabilities of Random Forest, the algorithm can be used to understand variable importance. Because many regression trees are being produced with different sets of input variables, the Random Forest algorithm learns and records the relative importance of the input variables in predicting the output. This capability is especially attractive as one of the objectives of this study is to understand the relative importance of explanatory variables in predicting street flooding, thus directing future investments in improving observational networks within the city.

3 Methods

3.1 Input Data Pre-processing

All of the raw input environmental data were aggregated to match the time scale of the flood reports. For all of the days on which no flood reports were made, and for storm events resulting in flood reports made only on one day, the data were aggregated to a daily time scale. For the storm events whose flood reports spanned multiple days, the data were aggregated across the days so that each storm event had only one set of average environmental conditions. For example, flood reports labeled "Hurricane Sandy" were recorded over three days, 2012-10-27, 2012-10-28, and 2012-10-29. The higher high tide taken for this event was the highest of the higher high tides of these three days, the average level of the surficial groundwater table was the average over these three days, and the total cumulative rainfall was the accumulated rainfall from the three days. The resulting dataset consisted of 2,171 records of average environmental conditions, mostly at a daily time scale, from September 2010 through October 2016. The aggregated environmental input variables are shown in Table 2.

Different approaches of aggregation were taken for the various environmental data. Four derivatives of the raw HRSD 15-minute rainfall data were included in the models as inputs: total cumulative rainfall, maximum hourly rainfall, maximum 15-minute rainfall, and cumulative rainfall in the previous three days. The different derivatives of the rainfall data were included to account for different types of storm events that may cause flooding. For example, during convective thunderstorms in the summertime, the maximum 15-minute rainfall would be high, but the total cumulative rainfall may be low. For nor'easters, the cumulative rainfall would be high while the maximum 15-minute rainfall may be low.

As with the 15-minute rainfall data, several tide-related variables were model inputs including high and low tides, and average tide level. In coastal cities, such as Norfolk, the timing of rainfall and the tide levels can have an effect on flooding. For example, if tide level is especially high when a large amount of rain falls, the stormwater oulets may be submerged. Such tailwater conditions do not provide sufficient head difference for gravity-driven storm drainage systems to function properly resulting in more flooding than if the tide were low and the same amount of rain fell. To account for such interactions between tide and rainfall, the tide level at the time of the maximum 15-minute rainfall and the tide level at the time of the maximum hourly rainfall were included as model inputs.

The environmental conditions data were averaged across all the stations that recorded the variable. For example, the "Daily cumulative rainfall" is the total cumulative rainfall averaged across all 11 rain gauges. This spatial averaging was done because for some of the stations there was a considerable amount of missing data over the six years of the study period. If the variables were not averaged across measuring stations, it would appear that the stations that had less missing data were more important which would make it more difficult to understand the importance of the actual environmental variables compared to the consistency of measurements at an individual station.

To reduce noise in the data, days on which little or no rainfall was recorded were not used in the modeling procedure. Of the 45 events for which flooding was reported, 42 had an average cumulative rainfall total of 0.25 mm or greater. Of the three events with less than 0.25 mm of rainfall, only one flooded location was reported for two of the events and only two flooded locations were reported for the third event. Given that very minor flooding was reported for days without rainfall, days with little to no rainfall (<0.25 mm of cumulative

Table 2. Input feature names and descriptions

Input Feature	Units	Source Organization	Abbreviation
Total cumulative rainfall	mm	airports, HRSD	RT
Maximum hourly rainfall	mm	HRSD	RHRMX
Maximum 15-minute rainfall	mm	HRSD	R15MX
Cumulative rainfall in previous three days	mm	HRSD	R3D
Average water table elevation	m above NAVD88	HRSD	GW_AV
Average tide level	m above MSL	NOAA	TD_AV
Tide level at time of maximum 15-minute rainfall	m above MSL	NOAA	TD_R15
Tide level at time of maximum hourly rainfall	m above MSL	NOAA	TD_RHR
High tide	m above MSL	NOAA	HT
Higher high tide	m above MSL	NOAA	HHT
Low tide	m above MSL	NOAA	LT
Lower low tide	m above MSL	NOAA	LLT
Average daily wind speed	km per hour	airports, HRSD, NOAA	AWND
Average daily wind direction	degrees	airports, HRSD, NOAA	AWDR
Average wind speed over 6-minutes	km per hour	airports, HRSD, NOAA	WSF6
Average wind direction over 6-minutes	degrees	airports, HRSD, NOAA	WDF6
Average maximum 2-minute wind gust over 6-minutes	km per hour	airports, HRSD, NOAA	WGF6
Average wind speed over 2-minutes	km per hour	airports, HRSD, NOAA	WSF2
Average wind direction over 2-minutes	degrees	airports, HRSD, NOAA	WDF2

rainfall) were considered in the model training and evaluation. This reduced the number of total records used to train and evaluate the model from 2,171 to 814.

3.2 Model Training and Evaluation

Model training and evaluation were performed using two independent, randomly selected partitions of the output and corresponding input data. In some studies, the dataset is split into three partitions, a training, evaluation, and validation set [*Tao et al.*, 2017], however, since a two-way split is common in this field [*Tien Bui et al.*, 2016; *Tehrany et al.*, 2013; *Solomatine and Price*, 2004] and the available dataset is of limited volume, the dataset was split into only two partitions. In the model training, the evaluation dataset was withheld and the models were fit to only the training data. By withholding the evaluation dataset in model training, the models can be evaluated using data not previously seen by the models, thus simulating actual use of the predictive models.

The R programming language (version 3.3.3) was used to partition the datasets, train the two models, and apply the models to the unseen, evaluation dataset [*R Core Team*, 2017]. The dataset of environmental conditions (the model input data) for storm events from September 2010 to October 2016 and the number of reported flood locations for each event (the model output data) were randomly divided into a training set (70%) and an evaluation set

(30%) [*Tien Bui et al.*, 2016; *Tehrany et al.*, 2013; *Solomatine and Price*, 2004] using the "caret" package in R [*Kuhn et al.*, 2016]. This package supports the stratified sampling of the datasets based on the distribution of the model output data, which included the 42 storm events for which flooding was reported and the 772 events for which no flooding was reported. By using stratified sampling, the distribution of the number of reported floods (the output variable) in both the training and evaluation datasets was similar to the distribution of the number of reported floods of the entire data set. To account for potential bias in the division of the data into training and evaluation sets, the random division was made and the models were trained 100 times independently for both models.

Since Poisson regression is a parametric model, the training of the Poisson regression model consisted of optimizing the model coefficients. The built-in "stats" package in R was used for the Poisson regression model. The training of the Random Forest consisted of training each of the individual regression trees making up the Random Forest. The 'randomForest' package (version 4.6.12) was used for the Random Forest model [*Liaw and Wiener*, 2002].

The Random Forest model has two main parameters, the number of trees per forest, and the number of random predictor variables each tree uses. A sensitivity study of these parameters was performed to determine their effect on model performance. To determine the model sensitivity to the number of trees per forest, the models were trained with the number of trees varying between 2 and 2,000 with the default number of variables per tree (i.e., one-third of the variables, or six in this case). The "tuneRF" function in the "randomForest" package was used to determine the optimum number of variables per regression tree in the Random Forest. This function changes the number of variables used in each regression tree to find the number of variables that minimizes the out-of-bag error within the Random Forest. The out-of-bag error is the prediction error when each input record is applied only to the portion of the regression trees which did not contain that input record in its training sample [*Breiman*, 2001]. Only the training data was used to determine the appropriate number of trees and variables per tree [*Xu et al.*, 2017].

Once trained, both the input training dataset and the input evaluation dataset, which was withheld in the model training, were used as input for the models. Applying the models to the input data produced a predicted number of reported floods for each array of input values. The predicted numbers of reported floods were compared with the known number of reported floods. Root mean squared error (RMSE) and mean absolute error (MAE) between the known and predicted number of flood reports were the two main metrics used to evaluate the models. Since each model was trained 100 times (once for each of the 100 random divisions into training and evaluation data), a distribution of predicted number of flood reports was produced in the model evaluation for each known number of reported floods. To describe these distributions, their standard deviations (std) were plotted and the average standard deviation of each model was reported.

In addition to RMSE, MAE, and std, the models' false negative and false positive predictions were also used to evaluate the models. False negative predictions occur when the predicted number flood reports is zero and the true number of flood reports is non-zero. False positive predictions occur when the true number of flood reports is zero and the predicted number of flood reports is non-zero. These terms are sometime referred to as Type I and Type II errors, respectively [*Beguería*, 2006]

False negative and false positive predictions are of particular interest to a decision maker. False negative predictions may jeopardize human safety and incur higher costs in recovery when a true positive prediction would have led to less costly, preventative measures. False positive predictions over time can erode trust in the warnings [*Basha et al.*, 2008]. For the Random Forest and Poisson regression models, the statistical characteristics (e.g., count, mean, standard deviation) of the false negative and false positive predictions were reported and compared. Since predictions were on a continuous scale and true flood reports were integers, the predictions were rounded to the nearest integer when calculating the false negative and false positive rates. For example, a prediction was considered false positive when the number of flood predictions was at least 0.5 (which would round to 1) and the true number of flood reports was zero.

4 Results and Discussion

4.1 Model Results

The results of the Poisson regression training and evaluation are shown in Table 3 and Figure 3. Predictions from the Poisson regression greater than 159 flood reports (the largest number of flood locations reported from any one event) were assumed to be outside a reasonable range and were therefore omitted. These predictions accounted for 0.3% of all predictions and 5.9% of the predictions greater than 0.5 flood report made in the evaluation phase.

Table 3. Summary of training and evaluation results for Poisson regression and Random Forest. All units are in number of flood reports.

		RMSE		MAE		std	
		Training	Evaluation	Training	Evaluation	Training	Evaluation
All days	Poisson	2.31	6.71	0.46	0.96	4.99	18.42
	RF	1.86	3.87	0.30	0.69	3.06	6.00
Non-zero flood days	Poisson	10.06	29.81	6.56	16.34	5.18	19.11
	RF	8.04	16.55	4.41	9.83	3.17	6.21

On the other hand, since the Random Forest model predictions are the average of each regression tree's prediction, the Random Forest predictions cannot exceed the range of training values. Therefore, none of the Random Forest predictions were omitted.

Figure 3 shows the predictions made by the Poisson regression model in the training and evaluation phases. The predictions made using the training data as input generally follow the one-to-one line, while the predictions made using the unseen, evaluation data as input are much more scattered. Additionally, for many of the values of true floods, the predicted number of floods in the evaluation phase had large standard deviations (mean of 18.42 flood reports) compared to the training phase (mean of 4.99 flood reports). For some values of true flood reports in the evaluation phase, the range of predictions was large even when the mean of the predictions was close to the true value. For example, the mean of the predictions when there were 31 true flood reports was 29, however, the predictions ranged from 4 to 91.

One explanation for the limited performance of the Poisson regression may be due to the target data not conforming to the assumptions used in the development of the Poisson regression model. One of the primary assumptions when using Poisson regression is that the variance and the mean of the output dataset are equal. In the flood reports dataset described in Section 2.2 and Figure 2, the mean was 1.2 flood reports, much lower than the variance, 108 flood reports, meaning that the data were overdispersed. A common method for handling overdispersed data in such cases is to use a modified version of Poisson regression called overdispersed Poisson. With overdispersed Poisson the assumption that the mean is equal to the variance is relaxed [*Gardner et al.*, 1995]. The overdispersed Poisson was tested as well but the results were very similar to the Poisson regression results.

From the sensitivity analysis of the Random Forest parameters, the number of trees per forest was 100. As seen in Figure 4 Random Forests with more than 100 regression trees saw minimal improvements in terms of RMSE, MAE, and std. The model was more sensitive to the number of variables per tree had a more significant impact. The number of variables per regression tree that performed the best in the optimization procedure was 16. Changing the number of variables per tree from six (the default) to 16 decreased the models RMSE by 23%.

The RMSE, MAE, and std of training and evaluation predictions from the Random Forest model are reported in Table 3. The RMSE was significantly higher in the evaluation phase compared to the training phase both when considering all of the events and when considering only events where floods were recorded. In both cases, the evaluation RMSE was about two times the training RMSE, suggesting that, like the Poisson regression, the model was overfit to the training data. Figure 5 shows the predicted number of flood reports made by the Random Forest model in the training and evaluation phases.

One reason for the overfitting seen in the models may be related to the imbalance of dataset. As would be expected, there are far more storm events where no flooding is reported, making the dataset imbalanced. The ratio of storm events for which some rain fell and zero flood reports were made to storm events on which some rain fell and at least one flood report was made is approximately 18:1. Another factor may be the relatively small nature of the dataset (less than 1,000 total records). *He and Garcia* [2009] noted that models trained with datasets that are both imbalanced and small are particularly prone to overfitting to the training data. As more data is collected and available for use in model training, it is expected that model overfitting would decrease.

Comparing the Poisson regression and Random Forest model, Random Forest performed better overall. In terms of RMSE and MAE, both models were nearly equal in the training phase, however, Random Forest performed significantly better in the evaluation phase. While both models showed signs of being overfit to the training data (i.e. a large drop in performance during the evaluation phase), the proportional difference in performance between training and evaluation in the Random Forest predictions was roughly two-thirds of that of the Poisson regression. A more significant difference in performance between Poisson regression and Random Forest was seen in the stability of the predictions in the evaluation phase as measured by the standard deviations of the predictions. Quantitatively, in the evaluation phase, the standard deviations of the Poisson predictions was more than three times that of the standard deviation of the Random Forest predictions. Visually, the difference is apparent when comparing the standard deviation bars in the evaluation predictions in Figures 3 and 5.

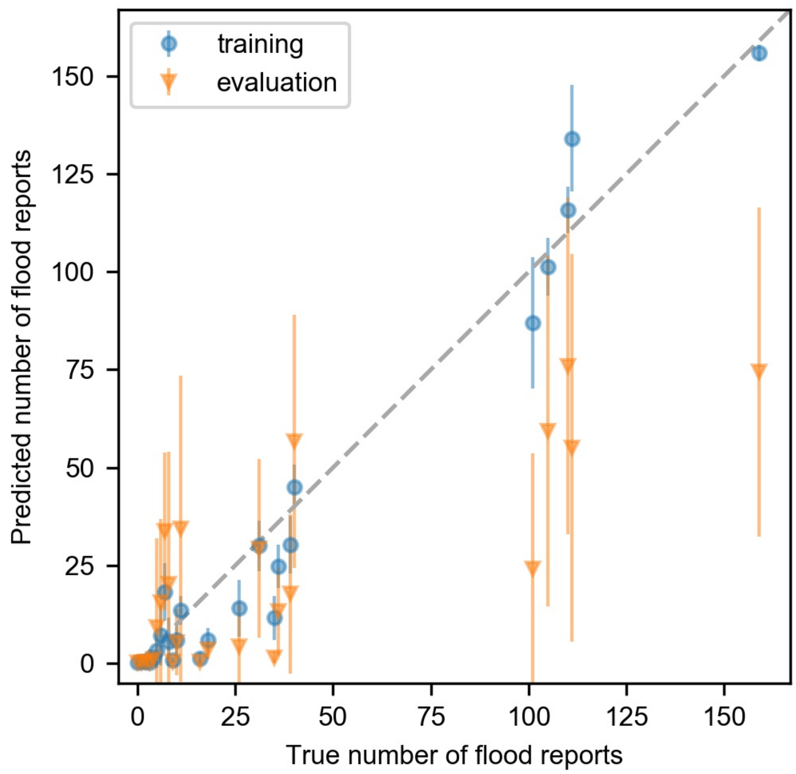

Figure 3. Model results for Poisson regression. Error lines represent the standard deviation of predictions.

It is important to note that model performance was measured using observed environmental conditions as model input. In practice, forecasted environmental conditions would be used as model input to predict flood severity. For example, rather than using rainfall and tide data observed at monitoring stations as input, rainfall forecasts from the High-Resolution Rapid Refresh (HRRR) model [*Smith et al.*, 2008] and NOAA's tide predictions [*NOAA*, 2017e] could be used as inputs. The use of uncertain forecast rainfall data are expected to increase the overall uncertainty of the model [*Collier*, 2007; *Bartholmes and Todini*, 2005].

4.2 False Negative and False Positive Predictions

Statistics summarizing false negative and false positive predictions in the evaluation phase are given in Table 4. Poisson regression had fewer but more variable and extreme false

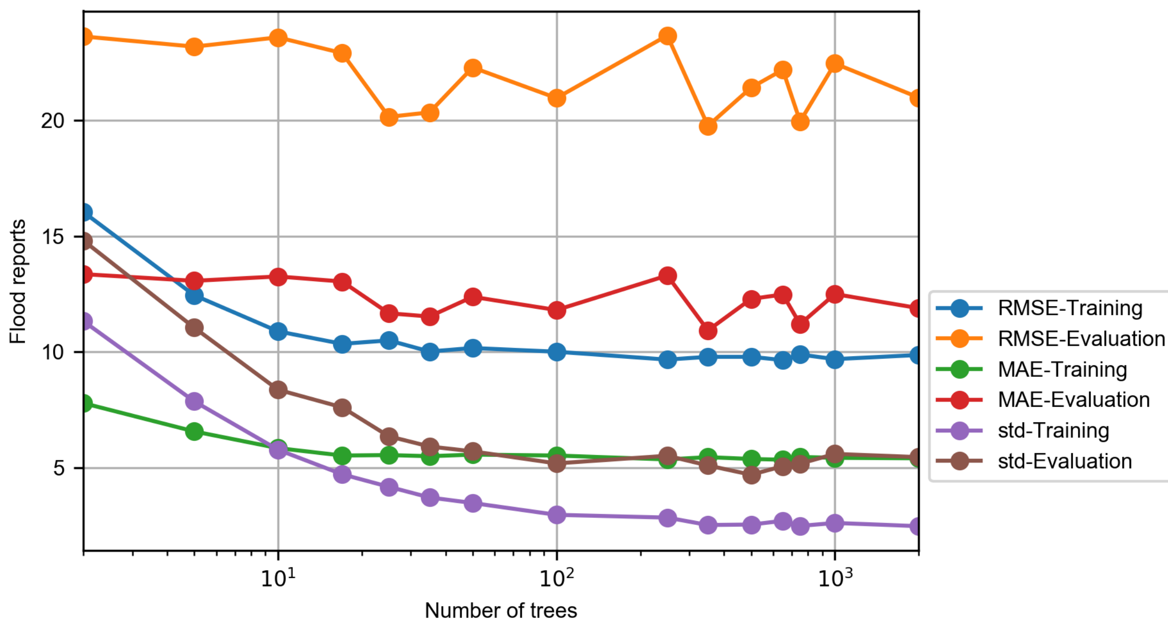

Figure 4. Random Forest model results with varying numbers of trees per model.

positive predictions compared to Random Forest. The false positive rate for Poisson regression was 4.39% compared to 7.09% for Random Forest, however, the standard deviation of the false positive predictions was much larger from the Poisson regression (8.55 flood reports compared to 3.18 flood reports). For both models, most of the false positive predictions were less than 1.17 flood reports, which would round down to one. The maximum false positive prediction was much greater in the Poisson regression compared to the Random Forest (122 compared to 36).

Compared to the false positive rates, both Poisson regression and Random Forest had much higher false negative rates (45.92% and 34.46% respectively). Importantly, compared to Random Forest, Poisson regression predicted false negatives when true flood reports were higher on average (mean of 6 flood reports compared to 3 flood reports). As with the false positives, the standard deviation of the false negatives predictions was higher for Poisson regression versus Random Forest, 11.55 flood reports compared to 2.30 flood reports. Similarly, the maximum true number of flood reports when a false negative prediction was made was much higher from Poisson regression (101 flood reports compared to 9 flood reports).

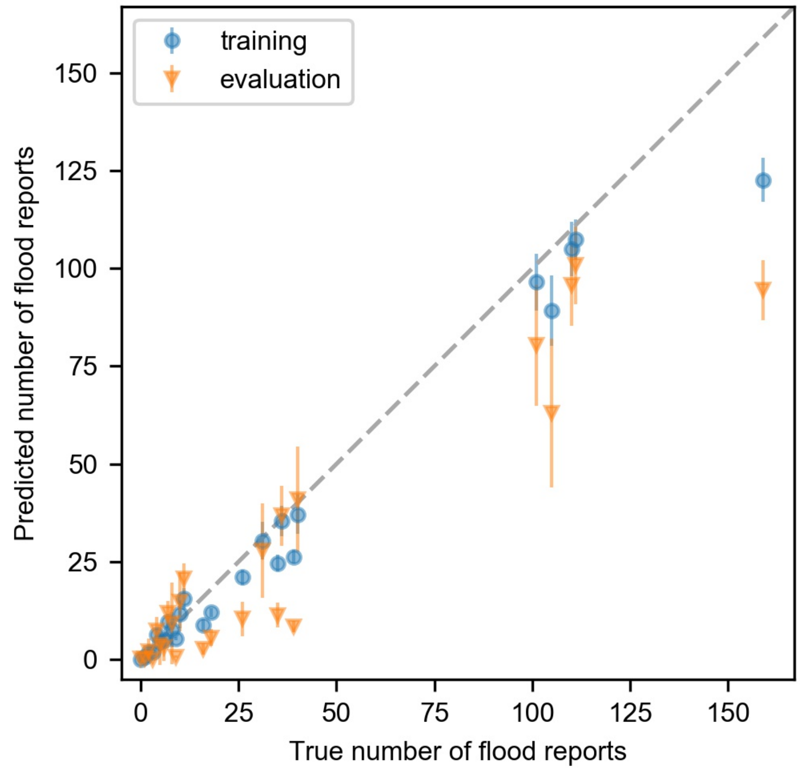

Figure 5. Model results for Random Forest. Error lines represent the standard deviation of predictions.

4.2.1 Variable Importance

Figure 6 shows the importance of each of the input variables as calculated from the Random Forest model in terms of the percent increase in mean squared error (MSE) when each of the variables is permuted individually. The total cumulative rainfall value was by far the most important of the variables. This was much more important than any of the other variables derived from the raw rainfall data, including the maximum hourly and maximum 15-minute rainfall values, suggesting that, in this record, large rainfall volumes caused more flooding than high rainfall intensities. The next three variables in terms of prediction importance were related to tide: low tide, lower low tide, and higher high tide.

The variable importance results shown in Figure 6 are supported by the raw data shown in Figure 7. The number of flood reports has clear positive relationship with the total cumulative rainfall. The same is true for low tide and lower low tide. The relationship is less clear for the maximum hourly rainfall and the maximum 15-minute rainfall, but according

Table 4. False positive and false negative statistics for Poisson regression and Random Forest. The statistics in the false positive columns describe the predicted flood reports greater than 0.5 when the true number of flood reports was zero. The statistics in the false negative columns describe the true non-zero flood reports when the predicted flood reports were zero.

	False Positives		False Negatives	
	Poisson	RF	Poisson	RF
rate (%)	4.39	7.09	45.92	34.46
count	1023.00	1653.00	524.00	428.00
25%	0.69	0.70	1.00	1.00
50%	1.10	1.17	2.00	2.00
75%	2.40	2.81	5.00	3.00
max	122.48	36.37	101.00	9.00
mean	3.41	2.47	5.77	2.64
std	8.55	3.18	11.55	2.30

to Figure 6, the Random Forest model was still able to glean some meaningful information from these variables. Interestingly, the tide level during the maximum 15-minute rainfall, has a clearer visual relationship with the number of flood reports compared to the maximum hourly, and maximum 15-minute rainfall values, but is considered less important by the Random Forest algorithm. One explanation for this is that the information provided by the tide level during the maximum 15-minute rainfall is already provided to the model, perhaps in a more useful form, from the low tide, lower low tide, and higher high tide variables.

The average height of the water table during a given event, surprisingly, did not add appreciable predictive power to the model. This may suggest that the surficial groundwater table did not impact flood severity in a significant way. However, the fact that it did not provide predictive power does not necessarily mean that the surficial groundwater table level did not contribute to flooding. For example, since the infiltration of rainfall causes the surficial groundwater table to rise, it is possible that the information provided by the rainfall data provides similar but clearer predictive power to the model compared to the surficial groundwater table level.

Although total cumulative rainfall is clearly the dominant predictor of flood severity in this dataset, it is commonly understood that other factors can have a significant impact on flooding in a coastal environment. For example, high tides alone can cause flooding in coastal cites [*Marfai et al.*, 2008]. When tidal information is omitted from the model inputs in this study, the RMSE of the Random Forest predictions increases by 4% and by 14% for the Poisson regression predictions. Thus, while rainfall is clearly the most important variable, tide levels and potentially other environmental variables cannot be ignored. It is anticipated that as sea levels rise, the importance of tide levels and water table level in predicting flooding will grow [*Hoover et al.*, 2016].

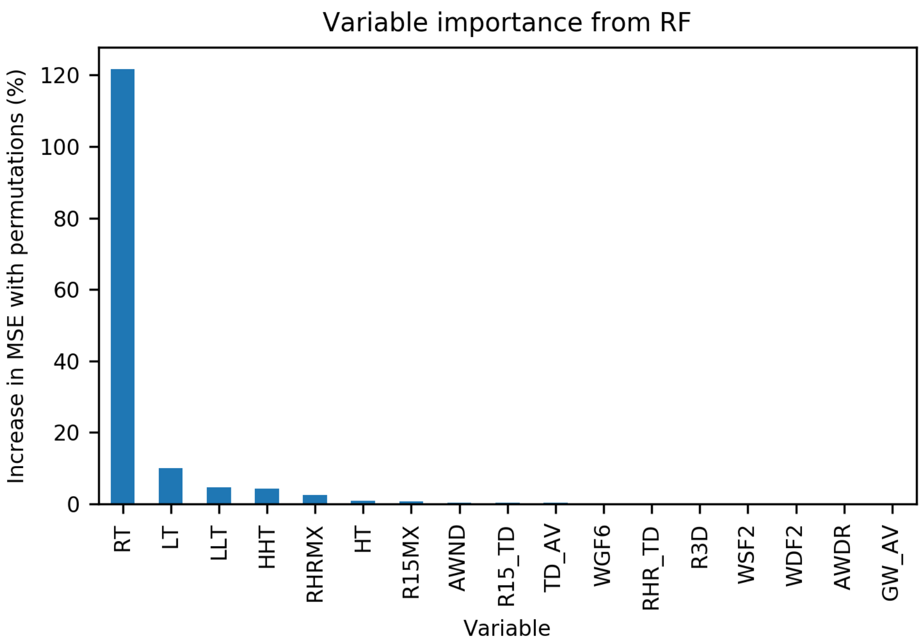

Figure 6. Importance of input variables

4.3 Potential Explanations for Model Limitations

A likely reason for the limited performance in both models is the limited amount of reported street floods used to train the models. The crowd-sourced flood report dataset used in this study is a unique and valuable dataset, but still a complete picture of flooding impacts is missing. Flood reports were made on only 42, or just over 5%, of the records used in the modeling. In addition, the number of flooding reports were distributed very unequally among the 42 storm events on which flood reports were made. More than 65% of the total flood reports were recorded from just six storm events (0.6% of the total storm events modeled). The rarity of storm events with any flood reports, and especially a large number of flood reports, makes it difficult for the model training. *Solomatine and Price* [2004] faced similar prob-

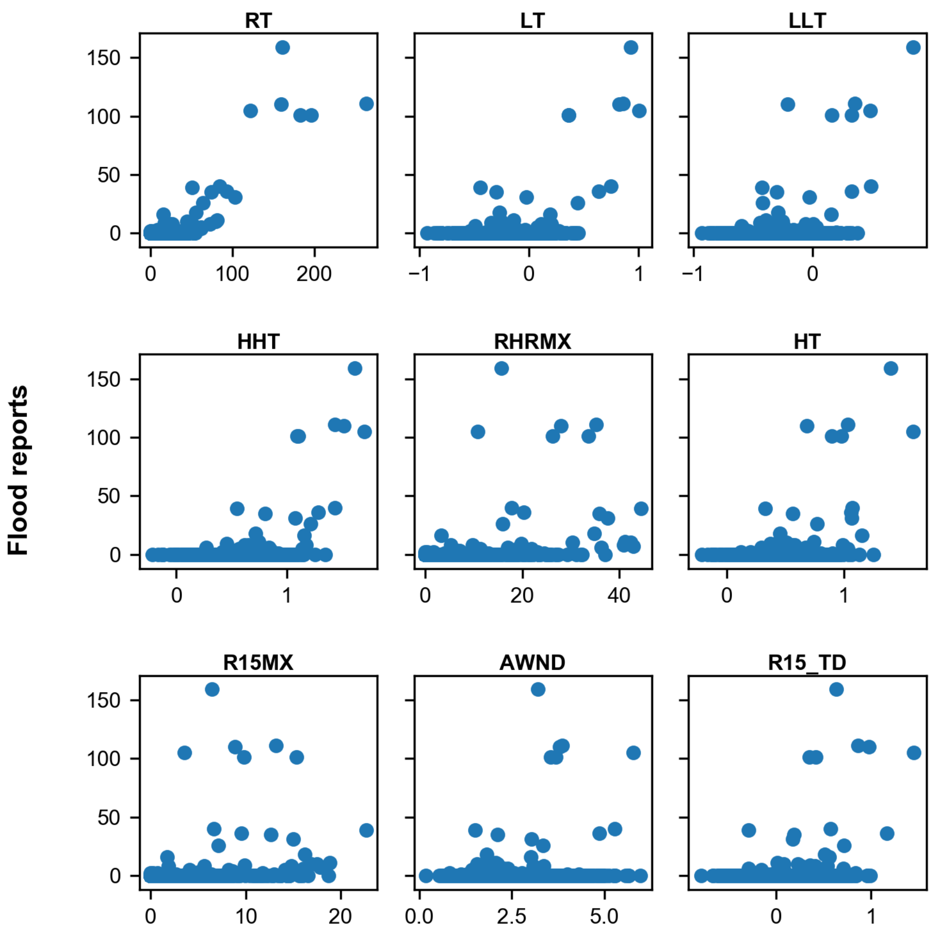

Figure 7. Flood reports against top nine predictor variables. Units for each variable are shown in Table 2

lems in training their machine learning model to accurately predict high peak flows which occurred rarely in their dataset.

The results also suggest that, compared to storm events with large volumes of rainfall which caused flooding, other types of storm events were not as well modeled by the data-driven models. Figure 9 shows the percent error of the Random Forest evaluation predictions for the 11 events with the top 10 number of reported floods (two events had 101 flood reports: Hurricane Nicole and an unnamed event occurring on 20 September 2016). Two unnamed events, heavy rain which occurred on 16 May 2014 (35 reported flood locations) and thunderstorms which occurred on 10 July 2014 (39 reported flood locations) have average percent error magnitudes larger than the rest. Both of these events had much lower cumulative rainfall and tide levels, the most important variables in the model (see Figure 7), but higher maximum hourly rainfall and relatively high maximum 15-minute rainfall values compared to the other high flooding events. Given this, it is possible that these events caused flash floods. The worse performance of the models at predicting the flooding severity from these two events may suggest that this type of flooding is not well represented in the training dataset. It is expected that the data-driven models would better predict such flood events with a larger, more complete dataset, containing more instances of similar flooding events. Additionally, with more training data, the model could be trained on specific subsets of flood events tailoring it to a type of flood event with specific characteristics (e.g., flash floods).

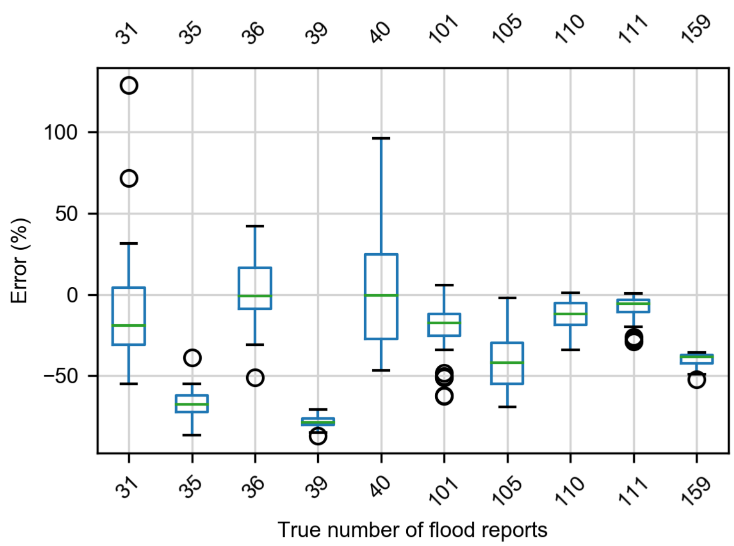

Figure 8. Top 10 flooding event percent error from Random Forest evaluation results

Besides the limited number of flooding events with which to train the models, bias present in the training data could have hampered model performance. Because the flooded locations were reported by individuals, there is an unknown amount of subjectivity and bias in the data as can be expected when using crowd-sourced data. Since the models are trained on data reported by individuals, one individual may influence the trained model disproportionally. Figure 10 shows the total number of flood reports made by each individual reporter, and the number of flood events for which each reporter recorded at least one flooded location during the period of record. The highest number of total reports made by one reporter was 158, 14% of the sum total reports from all 71 reporters. Therefore, the models in their training, are significantly influenced by this one reporter and can inherit, to some extent, his/her biases.

Another potential bias is in the under- or over-representation of different roadway types in the flooding record. Figure 11 shows the percentage of roadway length per VDOT roadway class in Norfolk and the percent of each roadway class at which flood reports were made (Table 5 gives the descriptions for each of the classes). From the figure, it is seen that although public local streets (class 6) account for the majority of the roadway length of the city (close to 60%), only 40% of the flooded streets reported were public local streets. Conversely, principal arterials (class 3) accounted for nearly 30% of the flooded street reports even though these streets make up less than 10% of Norfolk's total roadway length. This suggests that a flooded street less traveled and, therefore, less important to the overall connectivity of the city's street network, may have flooded but may not have been reported within the record with the same frequency as the more major roads.

A third example of bias may occur when unequal attention in reporting is given to certain geographic areas of the city or to certain storm events. One example of this bias is seen in the difference in reported floods between Hurricane Hermine and Hurricane Matthew which occurred only one month apart. For Hurricane Hermine, 22 flood reports, more than half of the total of 40 flood reports made as a result of Hurricane Hermine, came from one area in downtown Norfolk. In contrast, for Hurricane Matthew, which produced more than three times as much total cumulative rainfall on average than Hurricane Hermine (264 mm compared to 84 mm) and was at least comparable in terms of tide, water table height, and wind conditions, only six flood reports were made from the same area. It is unlikely that the actual flooding caused by Hurricane Matthew, a much larger storm, was in fact a quarter in severity, but more likely that there were significant differences in reporting between the two events.

4.4 Increase in Street Flood Reports

Flooding reports and events have become more frequent over the period of record (September 2010 to October 2016). The number of flooded street reports has increased year to year in the past four years and overall in the past seven years (see Figure 12). More than twice as many floods were reported in 2016 compared to 2014. Very few flood reports were made in 2013 compared to the other years of record. This can be explained, at least in part, because 2013 was an exceptionally mild hurricane season, the first since 1994 without any major hurricanes. The only storm event in 2013 reported to have caused flooding was an unnamed heavy rain event.

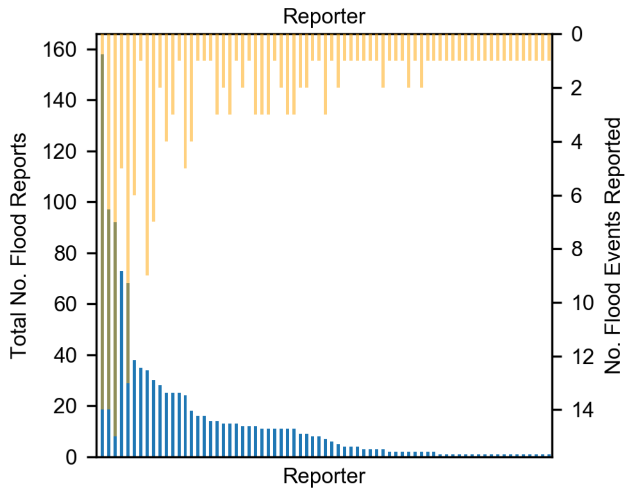

Figure 9. Number of total reports made and events reported per reporter

Table 5. VDOT roadway class codes and descriptions

VDOT Road Class Code	Description
1	Interstate
2	Tunnel Roads and other VDOT owned
3	Principal Arterials
4	Minor Arterials
5	Collectors
6	Local Streets- Public
7	Local Streets- Private
8	Miscellaneous
9	Base Roads
10	Public Alleys

The overall increase in the number of flood reports over the period of record was due primarily to an increase in the number of storm events from which flood reports were made rather than an increase in the number of reports per storm event. In the years 2010-2013, five total storm events were reported to have caused flooding, while in 2014 alone, 16 events were reported to have caused flooding and at least 10 storm events per year resulted in reports of flooded streets in 2015 and 2016 (see Figure 12). In contrast to the storm events reported in 2010-2013, most of the storm events reported to have caused flooding in the years 2014-2016 were smaller, unnamed storm events. In each of the years 2010, 2011, and 2012, there was a storm event for which more than 100 flood reports were made, each of them named, major hurricanes (Nicole, Irene, and Sandy, respectively). In 2014, on the other hand, of the 16

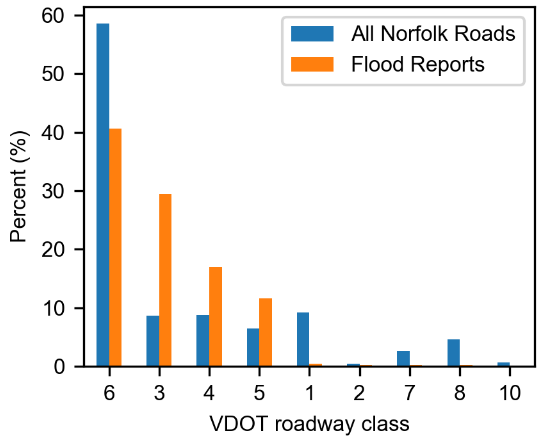

Figure 10. Percentage of total roadway length and percentage of reported floods per VDOT roadway class in Norfolk, VA

storm events reported to have caused flooding, the maximum number of flood reports for an individual storm event was 39, and only one named hurricane was reported to have caused flooding, Hurricane Arthur.

The increase in flood reports due to smaller storms from 2014-2016 may suggest that the City of Norfolk is becoming more susceptible to flooding from less extreme storm events (i.e. storm events which are not hurricanes). It is noted that most of the flood reports were made by staff of the City which operates with limited resources. In reality, the number of flood locations resulting from the recorded storm events could be larger than what was reported, including street floods that may have occurred from less extreme events in 2010-2013. It is possible that the increase in flood reports over the period of record may simply be due to an increase in attention given to street flooding and resources to street flood reporting in Norfolk, rather than an increase in actual flooding. An increase in attention and resources allocated to street flooding and street flood reporting from the City of Norfolk however, may still suggest that flooding problems are worsening.

Sweet and Park [2014] noted an increase of nuisance level tidal flooding in Norfolk, VA from 1.2 days per year in the years 1956-1960 to 7.4 days per year in the years 2006-2010. *Sweet and Park* [2014] also predicted that the amount of flooding will increase with time due to predicted sea level rise. This is in agreement with the increase of flood reports seen in the present study. As sea level rises and climate changes, it may be necessary to incorporate a mechanism to account for this change into the data-driven models. This change be referred to as concept drift [*Gama et al.*, 2014; *Gama and Castillo*, 2006] and mechanisms for accounting for it would be especially useful when using Random Forest, which as noted above, cannot exceed the range of training data.

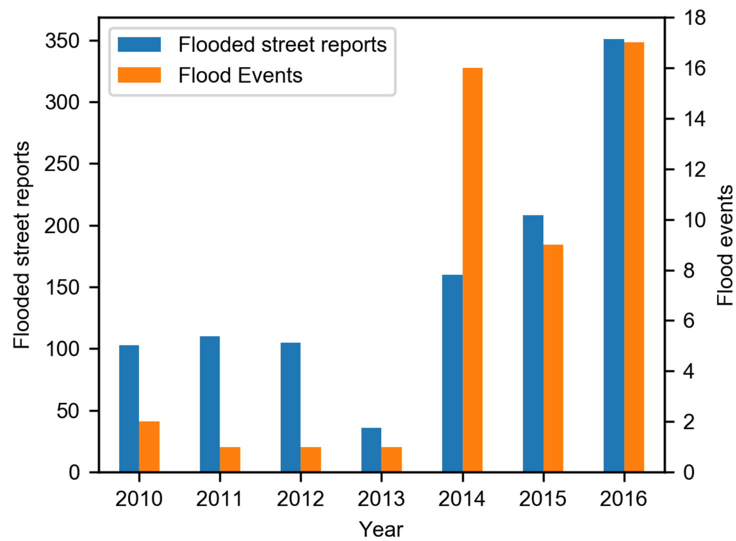

Figure 11. Flood events and in Norfolk, VA Sep. 2010 - Oct. 2016

5 Conclusions

Two data-driven models, Poisson regression and Random Forest were trained to predict flood severity for a given set of environmental conditions (rainfall, tide levels, groundwater levels, and wind conditions) using quality-controlled, crowd-sourced street flooding reports as a proxy output variable. The data used for training and evaluating the models was from Norfolk, Virginia USA. The Random Forest model performed better overall compared to Poisson regression in the evaluation phase (root mean squared error of 3.87 compared to 6.71 flood reports, mean absolute error of 0.69 compared to 0.96 flood reports) with less variance (standard deviation of 6.00 compared to 18.42 flood reports). The most important variable in predicting model output in the Random Forest model was by far total cumulative rainfall followed by low tide and lower low tide.

The quality-controlled crowd-sourced record provided by the City of Norfolk, despite limited coverage spatially and from storm to storm, provided an uncommonly detailed flood record. In the record, flooding at individual intersections and streets was recorded for many events over an extended period of time. Using this as training data, the models demonstrated in this paper could give city workers a reasonable estimate of flooding severity based on forecasted environmental conditions. This is a first step in the long-term goal of spatially and temporally detailed urban flood predictions to assist city managers in real-time flood adaptation measures such as traffic management. This also demonstrates one way that crowd-sourced data, despite its limitations, can provide useful information to flood prediction models.

A main limiting factor in building accurate models is the quantity and quality of the record of flooding used to train the model. Given the bias present in the training dataset, predictions were necessarily lumped spatially (predictions were made at the city scale) and temporally (predictions were made at a event time scale). While other work has raised the need for accurate and dense measurements of rainfall [*Sadler et al.*, 2017; *Hill et al.*, 2014], the primary input to flood models, the results of this work highlight the need for more accurate and complete record of flooding data including depth and duration of flood occurrences. Such data is needed to adequately train flood models with enough spatial and temporal detail to help make street-level, real-time operational decisions.

Given more complete and objective flood occurrence data, it is likely that a data-driven model such as the ones demonstrated in this paper, could predict street flooding with much greater precision. The need for a more complete flood record data may be filled with a street-level sensor network, eliminating human subjectivity. Such a network is currently being piloted to record water levels at commonly flooded intersections in Norfolk. The detailed data from this network could be used to further improve predictions from models such as those demonstrated in this paper. Crowd-sourced data such as flood reports made from cellular devices could also be useful. Although the subjectivity in publicly crowd-sourced data would likely be similar to the dataset used in the paper, a wider number of reporters would presumably mitigate the subjectivity to some degree.

Data Availability

Data used for the analysis in this study can be found on HydroShare:

https://www.hydroshare.org/resource/9db60cf6c8394a0fa24777c8b9363a9b/.

Acknowledgments

We gratefully acknowledge the Mid-Atlantic Transportation Sustainability Center University Transportation Center for funding of this work. We also thank the Hampton Roads Sanitation District for access to their data. Lastly we thank the City of Norfolk for providing the reported flood data.

References

Adamovic, M., F. Branger, I. Braud, and S. Kralisch (2016), Development of a data-driven semi-distributed hydrological model for regional scale catchments prone to Mediterranean flash floods, *Journal of Hydrology*, *541*, 173–189, doi:10.1016/j.jhydrol.2016.03.032.

Bartholmes, J., and E. Todini (2005), Coupling meteorological and hydrological models for flood forecasting, *Hydrology and Earth System Sciences*, *9*(4), 333–346, doi:10.5194/hess-9-333-2005.

Basha, E. A., S. Ravela, and D. Rus (2008), Model-based monitoring for early warning flood detection, in *Proceedings of the 6th ACM conference on Embedded network sensor systems - SenSys '08*, p. 295, ACM Press, New York, New York, USA, doi:10.1145/1460412.1460442.

Bates, P. D., R. J. Dawson, J. W. Hall, M. S. Horritt, R. J. Nicholls, J. Wicks, and Mohamed Ahmed Ali Mohamed Hassan (2005), Simplified two-dimensional numerical modelling of coastal flooding and example applications, *Coastal Engineering*, *52*(9), 793–810, doi:10.1016/j.coastaleng.2005.06.001.

Beguería, S. (2006), Validation and Evaluation of Predictive Models in Hazard Assessment and Risk Management, *Natural Hazards*, *37*(3), 315–329, doi:10.1007/s11069-005-5182-6.

Booker, D. J., and R. A. Woods (2014), Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments, *Journal of Hydrology*, *508*, 227–239, doi:10.1016/j.jhydrol.2013.11.007.

Breiman, L. (2001), Random Forests, *Machine Learning*, *45*(1), 5–32, doi:10.1023/A:1010933404324.

Breiman, L., J. Friedman, R. Olshen, and C. Stone (1984), *Classification and Regression Trees*, Wadsworth, Belmont, CA.

Broder, J. M. (2009), Climate change seen as threat to US security, *The New York Times*, *9*, 1.

Cameron, A. C., and P. K. Trivedi (1998), *Regression analysis of count data*, vol. 53, Cambridge university press.

Caviedes-Voullième, D., P. García-Navarro, and J. Murillo (2012), Influence of mesh structure on 2D full shallow water equations and SCS Curve Number simulation of rainfall/runoff events, *Journal of Hydrology*, *448-449*, 39–59, doi:10.1016/j.jhydrol.2012.04.006.

Church, J. A., J. M. Gregory, P. Huybrechts, M. Kuhn, K. Lambeck, M. T. Nhuan, D. Qin, and P. L. Woodworth (2001), Changes in sea level, in *in: JT Houghton, Y. Ding, DJ Griggs, M. Noguer, PJ Van der Linden, X. Dai, K. Maskell, and CA Johnson (eds.): Climate Change 2001: The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel*, pp. 639–694.

Collier, C. G. (2007), Flash flood forecasting: What are the limits of predictability?, *Quarterly Journal of the Royal Meteorological Society*, *133*(622), 3–23, doi:10.1002/qj.29.

Coxe, S., S. G. West, and L. S. Aiken (2009), The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives, *Journal of Personality Assessment*, *91*(2), 121–136, doi:10.1080/00223890802634175.

D'Unger, A., K. C. Land, P. L. McCall, and D. S. Nagin (1998), How Many Latent Classes of Delinquent/ Criminal Careers? Results from Mixed Poisson Regression Analyses, *American Journal of Sociology*, *103*(6), 1593–1630, doi:10.1086/231402.

Ezer, T., and L. P. Atkinson (2014), Accelerated flooding along the U.S. East Coast: On the impact of sea-level rise, tides, storms, the Gulf Stream, and the North Atlantic Oscillations, *Earth's Future*, *2*(8), 362–382, doi:10.1002/2014EF000252.

Fears, D. (2012), Built on sinking ground, Norfolk tries to hold back tide amid sea-level rise.

Frome, E. L., and H. Checkoway (1985), USE OF POISSON REGRESSION MODELS IN ESTIMATING INCIDENCE RATES AND RATIOS, *American Journal of Epidemiology*, *121*(2), 309–323, doi:10.1093/oxfordjournals.aje.a114001.

Gaitan, S., N. van de Giesen, and J. ten Veldhuis (2016), Can urban pluvial flooding be predicted by open spatial data and weather data?, *Environmental Modelling & Software*, *85*, 156–171, doi:10.1016/J.ENVSOFT.2016.08.007.

Gallien, T. W., B. F. Sanders, and R. E. Flick (2014), Urban coastal flood prediction: Integrating wave overtopping, flood defenses and drainage, *Coastal Engineering*, *91*, 18–28, doi:10.1016/j.coastaleng.2014.04.007.

Gama, J., and G. Castillo (2006), Learning with Local Drift Detection, pp. 42–55, Springer, Berlin, Heidelberg, doi:10.1007/11811305_4.

Gama, J., I. ŽliobaitÄŮ, A. Bifet, M. Pechenizkiy, and A. Bouchachia (2014), A survey on concept drift adaptation, *ACM Computing Surveys*, *46*(4), 1–37, doi:10.1145/2523813.

Gardner, W., E. P. Mulvey, and E. C. Shaw (1995), Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models., *Psychological bulletin*, *118*(3), 392.

Godschalk, D. R. (2003), Urban Hazard Mitigation: Creating Resilient Cities, *Natural Hazards Review*, *4*(3), 136–143, doi:10.1061/(ASCE)1527-6988(2003)4:3(136).

He, H., and E. A. Garcia (2009), Learning from imbalanced data, *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284, doi:10.1109/TKDE.2008.239.

Hill, D., B. Kerkez, A. Rasekh, A. Ostfeld, B. Minsker, and M. K. Banks (2014), Sensing and Cyberinfrastructure for Smarter Water Management: The Promise and Challenge of Ubiquity, *Journal of Water Resources Planning and Management*, *140*(7), 01814,002, doi: 10.1061/(ASCE)WR.1943-5452.0000449.

Hoover, D. J., K. O. Odigie, P. W. Swarzenski, and P. Barnard (2016), Sea-level rise and coastal groundwater inundation and shoaling at select sites in California, USA, *Journal of Hydrology: Regional Studies*, doi:10.1016/j.ejrh.2015.12.055.

Horsburgh, J. S., D. G. Tarboton, M. Piasecki, D. R. Maidment, I. Zaslavsky, D. Valentine, and T. Whitenack (2009), An integrated system for publishing environmental observations data, *Environmental Modelling & Software*, *24*(8), 879–888.

Hunter, N., P. Bates, S. Neelz, G. Pender, I. Villanueva, N. Wright, D. Liang, R. Falconer, B. Lin, S. Waller, A. Crossley, and D. Mason (2008), Benchmarking 2D hydraulic models for urban flood simulations, *Proceedings of the institution of civil engineers: water management*, *161*(1), 13—-30.

Ireland, G., M. Volpi, and G. Petropoulos (2015), Examining the Capability of Supervised Machine Learning Classifiers in Extracting Flooded Areas from Landsat TM Imagery: A Case Study from a Mediterranean Flood, *Remote Sensing*, *7*(3), 3372–3399, doi:10.3390/rs70303372.

Kerkez, B., C. Gruden, M. Lewis, L. Montestruque, M. Quigley, B. Wong, A. Bedig, R. Kertesz, T. Braun, O. Cadwalader, A. Poresky, and C. Pak (2016), Smarter Stormwater Systems, *Environmental Science & Technology*, *50*(14), 7267–7273, doi:10.1021/acs.est.5b05870.

Kleinosky, L. R., B. Yarnal, and A. Fisher (2006), Vulnerability of Hampton Roads, Virginia to Storm-Surge Flooding and Sea-Level Rise, *Natural Hazards*, *40*(1), 43–70, doi: 10.1007/s11069-006-0004-z.

Kuhn, M., J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, and B. Kenkel (2016), caret: Classification and Regression Training.

Lambert, D. (1992), Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing, *Technometrics*, *34*(1), 1, doi:10.2307/1269547.

Le Coz, J., A. Patalano, D. Collins, N. F. Guillen, C. M. Garcia, G. M. Smart, J. Bind, A. Chiaverini, R. Le Boursicaud, G. Dramais, and I. Braud (2016), Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand, *Journal of Hydrology*, *541*, 766–777, doi:10.1016/J.JHYDROL.2016.07.036.

Leandro, J., A. S. Chen, S. Djordjević, and D. A. Savić (2009), Comparison of 1D/1D and 1D/2D Coupled (Sewer/Surface) Hydraulic Models for Urban Flood Simulation, *Journal of Hydraulic Engineering*, *135*(6), 495–504, doi:10.1061/(ASCE)HY.1943-7900.0000037.

Liaw, A., and M. Wiener (2002), Classification and Regression by randomForest, *R News*, *2*(3), 18–22.

Loh, W.-Y. (2011), Classification and regression trees, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 14–23, doi:10.1002/widm.8.

Loos, M., and H. Elsenbeer (2011), Topographic controls on overland flow generation in a forest âĂŞ An ensemble tree approach, *Journal of Hydrology*, *409*(1-2), 94–103, doi:10.1016/j.jhydrol.2011.08.002.

Marfai, M. A., L. King, J. Sartohadi, S. Sudrajat, S. R. Budiani, and F. Yulianto (2008), The impact of tidal flooding on a coastal community in Semarang, Indonesia, *Environmentalist*, *28*(3), 237–248, doi:10.1007/s10669-007-9134-4.

Mark, O., S. Weesakul, C. Apirumanekul, S. B. Aroonnet, and S. Djordjevic (2004), Potential and limitations of 1D modelling of urban flooding, *Journal of Hydrology*, *299*(3-4), 284–299, doi:10.1016/j.jhydrol.2004.08.014.

Mignot, E., A. Paquier, and S. Haider (2006), Modeling floods in a dense urban area using 2D shallow water equations, *Journal of Hydrology*, *327*(1-2), 186–199, doi:10.1016/j.jhydrol.2005.11.026.

Moftakhari, H. R., A. AghaKouchak, B. F. Sanders, D. L. Feldman, W. Sweet, R. A. Matthew, and A. Luke (2015), Increased nuisance flooding along the coasts of the United

States due to sea level rise: Past and future, *Geophysical Research Letters*, *42*(22), 9846–9852, doi:10.1002/2015GL066072.

Morrison, J. E., and J. A. Smith (2002), Stochastic modeling of flood peaks using the generalized extreme value distribution, *Water Resources Research*, *38*(12), 41–1–41–12, doi: 10.1029/2001WR000502.

Murphy, K. P. (2012), *Machine learning: a probabilistic perspective*, MIT press.

Mutanga, O., E. Adam, and M. A. Cho (2012), High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm, *International Journal of Applied Earth Observation and Geoinformation*, *18*, 399–406, doi: 10.1016/j.jag.2012.03.012.

Naghibi, S. A., D. D. Moghaddam, B. Kalantar, B. Pradhan, and O. Kisi (2017), A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping, *Journal of Hydrology*, *548*, 471–483, doi:10.1016/j.jhydrol.2017.03.020.

NOAA (2017a), Sewells Point - Station Home Page - NOAA Tides & Currents.

NOAA (2017b), Money Point - Station Home Page - NOAA Tides & Currents.

NOAA (2017c), Daily Summaries Station Details: NORFOLK INTERNATIONAL AIRPORT, VA US, GHCND:USW00013737 | Climate Data Online (CDO) | National Climatic Data Center (NCDC).

NOAA (2017d), Daily Summaries Station Details: NORFOLK NAS, VA US, GHCND:USW00013750 | Climate Data Online (CDO) | National Climatic Data Center (NCDC).

NOAA (2017e), NOAA Tide Predictions Users Guide.

Osgood, D. W. (2000), Poisson-Based Regression Analysis of Aggregate Crime Rates, *Journal of Quantitative Criminology*, *16*(1), 21–43, doi:10.1023/A:1007521427059.

Povak, N. A., P. F. Hessburg, T. C. McDonnell, K. M. Reynolds, T. J. Sullivan, R. B. Salter, and B. J. Cosby (2014), Machine learning and linear regression models to predict catchment-level base cation weathering rates across the southern Appalachian Mountain region, USA, *Water Resources Research*, *50*(4), 2798–2814, doi:10.1002/2013WR014203.

R Core Team (2017), R: A Language and Environment for Statistical Computing.

Ray, R. D., and G. Foster (2016), Future nuisance flooding at Boston caused by astronomical tides alone, *Earth's Future*, doi:10.1002/2016EF000423.

Sadler, J. M., J. L. Goodall, and M. M. Morsy (2017), Effect of Rain Gauge Proximity on Rainfall Estimation for Problematic Urban Coastal Watersheds in Virginia Beach, Virginia, *Journal of Hydrologic Engineering*, *22*(9), 04017,036, doi:10.1061/(ASCE)HE.1943-5584.0001563.

Sahoo, S., T. A. Russo, J. Elliott, and I. Foster (2017), Machine learning algorithms for modeling groundwater level changes in agricultural regions of the U.S., *Water Resources Research*, *53*(5), 3878–3895, doi:10.1002/2016WR019933.

Smith, R. A., P. D. Bates, and C. Hayes (2011), Evaluation of a coastal flood inundation model using hard and soft data, *Environmental Modelling & Software*, *30*, 35–46, doi:10.1016/j.envsoft.2011.11.008.

Smith, T. L., S. G. Benjamin, J. M. Brown, S. Weygandt, T. Smirnova, and B. Schwartz (2008), CONVECTION FORECASTS FROM THE HOURLY UPDATED, 3-KM HIGH RESOLUTION RAPID REFRESH (HRRR) MODEL, in *Preprints, 24th Conference on Severe Local Storms*, Savannah, GA.

Solomatine, D. P., and A. Ostfeld (2008), Data-driven modelling: some past experiences and new approaches, *Journal of Hydroinformatics*, *10*(1), 3, doi:10.2166/hydro.2008.015.

Solomatine, D. P., and R. K. Price (2004), Innovative approaches to flood forecasting using data driven and hybrid modelling, in *Proc. 6th International Conference on Hydroinformatics*.

Svetnik, V., A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston (2003), Random forest: a classification and regression tool for compound classification and QSAR modeling, *Journal of chemical information and computer sciences*, *43*(6), 1947–1958.

Sweet, W. V., and J. Park (2014), From the extreme to the mean: Acceleration and tipping points of coastal inundation from sea level rise, *Earth's Future*, *2*(12), 579–600, doi:10.1002/2014EF000272.

Tao, Y., T. Yang, M. Faridzad, L. Jiang, X. He, and X. Zhang (2017), Non-stationary bias correction of monthly CMIP5 temperature projections over China using a residual-based bagging tree model, doi:10.1002/joc.5188.

Tehrany, M. S., B. Pradhan, and M. N. Jebur (2013), Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS, *Journal of Hydrology*, *504*, 69–79, doi:10.1016/j.jhydrol.2013.09.034.

The Port of Virginia (2016), Hampton Roads Harbor, 2016 Trade Overview.

Tien Bui, D., B. Pradhan, H. Nampak, Q.-T. Bui, Q.-A. Tran, and Q.-P. Nguyen (2016), Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using GIS, *Journal of Hydrology*, *540*, 317–330, doi:10.1016/j.jhydrol.2016.06.027.

Turner, R. E. (2004), Coastal wetland subsidence arising from local hydrologic manipulations, *Estuaries*, *27*(2), 265–272, doi:10.1007/bf02803383.

Vermeer, M., and S. Rahmstorf (2009), Global sea level linked to global temperature, *Proceedings of the National Academy of Sciences of the United States of America*, *106*(51), 21,527–21,532, doi:10.1073/pnas.0907765106.

Viglione, A., G. Di Baldassarre, L. Brandimarte, L. Kuil, G. Carr, J. L. Salinas, A. Scolobig, and G. Blöschl (2014), Insights from socio-hydrology modelling on dealing with flood risk âĂŞ Roles of collective memory, risk-taking attitude and trust, *Journal of Hydrology*, *518*, 71–82, doi:10.1016/j.jhydrol.2014.01.018.

Wang, W.-C., K.-W. Chau, C.-T. Cheng, and L. Qiu (2009), A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series, *Journal of Hydrology*, *374*(3), 294–306, doi:10.1016/j.jhydrol.2009.06.019.

Wang, Z., C. Lai, X. Chen, B. Yang, S. Zhao, and X. Bai (2015), Flood hazard risk assessment model based on random forest, *Journal of Hydrology*, *527*, 1130–1141, doi:10.1016/j.jhydrol.2015.06.008.

Xu, T., A. J. Valocchi, M. Ye, and F. Liang (2017), Quantifying model structural error: Efficient Bayesian calibration of a regional groundwater flow model using surrogates and a data-driven error model, *Water Resources Research*, doi:10.1002/2016WR019831.

Yang, T., X. Gao, S. Sorooshian, and X. Li (2016), Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme, *Water Resources Research*, *52*(3), 1626–1651, doi:10.1002/2015WR017394.

Yang, T., A. A. Asanjan, M. Faridzad, N. Hayatbini, X. Gao, and S. Sorooshian (2017a), An enhanced artificial neural network with a shuffled complex evolutionary global optimization with principal component analysis, *Information Sciences*, *418-419*, 302–316, doi:10.1016/j.ins.2017.08.003.

Yang, T., A. A. Asanjan, E. Welles, X. Gao, S. Sorooshian, and X. Liu (2017b), Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information, *Water Resources Research*, *53*(4), 2786–2812, doi:10.1002/2017WR020482.

Zevenbergen, C., W. Veerbeek, B. Gersonius, and S. Van Herk (2008), Challenges in urban flood management: travelling across spatial and temporal scales, *Journal of Flood Risk Management*, *1*(2), 81–88, doi:10.1111/j.1753-318X.2008.00010.x.